(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) Publication number: 0 339 891 A2

(12) EUROPEAN PATENT APPLICATION

(21) Application number: 89304017.0

(51) Int. Cl.4: G10L 3/00

(22) Date of filing: 21.04.89

(30) Priority: 23.04.88 JP 101173/88

(43) Date of publication of application:
02.11.89 Bulletin 89/44

(84) Designated Contracting States:
AT BE CH DE FR GB IT LI NL

(71) Applicant: CANON KABUSHIKI KAISHA
3-30-2 Shimomaruko Ohta-ku
Tokyo 146(JP)

(72) Inventor: Miyamae, Koichi
Canon Daiichi Honatsugi-ryo 6-29 Mizuhiki
2-chome
Atsugi-shi Kanagawa-ken(JP)
Inventor: Omata, Satoshi
1-5-101 Narusegaoka 1-chome
Machida-shi Tokyo(JP)

(74) Representative: Beresford, Keith Denis Lewis
et al
BERESFORD & Co. 2-5 Warwick Court High
Holborn
London WC1R 5DJ(GB)

(54) Speech processing apparatus.

(57) A speech processing apparatus of the present invention enables processor elements (403a to 403r) each comprising at least one nonlinear oscillator circuit (621) to be used as band pass filters by using the entrainment taking place in each of the processor elements, whereby the speech of a particular talker in the speech of a plurality of talkers can be recognized.

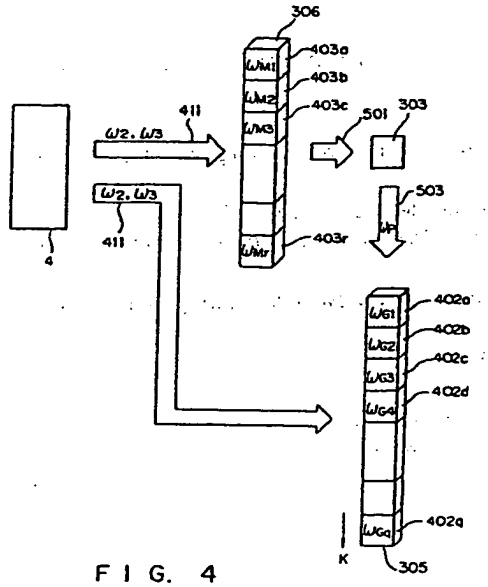

FIG. 4

EP 0 339 891 A2

input means for inputting speech from a plurality of talkers and outputting aural signals; a plurality of speech collation processor elements for performing speech collation using the aural signals input, each of the processor elements comprising at least one non-linear oscillator circuit which is designed to bring about the entrainment effect at the first frequency peculiar to the speech of a particular talker; a detection means for detecting the entrained state of each of the processor elements; and an extraction means for extracting the aural signals of a particular talker from the aural signals input therein when it receives the output from the detection means on the basis of the frequency of oscillations of the output signal of the processor element entrained.

It is another object of the present invention to provide a speech processing apparatus which is capable of recognizing at high speed constituent talkers of the conversation from the aural signals containing the speech of a plurality of talkers.

In order to achieve this object, the speech processing apparatus of the present invention is a speech processing apparatus which serves to specify the constituent talkers of the conversion input from a plurality of specified talkers and which comprises an input means for inputting conversational speech and outputting aural signals; a plurality of speech collation processor elements for performing speech collation using the aural signals input therein, each of the processor elements comprising at least one non-linear oscillator circuit which is designed to bring about the entrainment effect at the first frequency peculiar to the speech of a particular talker; and a detection means for detecting the entrained state of each of the processor elements.

It is a further object of the present invention to provide a speech processing system which is capable of performing as a whole speech information processing of a particular talker at high speed by extracting at high speed the speech of at least one particular talker from the aural signals containing the speech of a plurality of talkers and performing information processing such as speech recognition processing and so forth, e.g., word recognition and so on, of the aural signals extracted.

In order to achieve this object, the speech processing system of the present invention comprises an input means for inputting the speech from a plurality of talkers and outputting aural signals; a plurality of speech collation processor elements for performing speech collation of the aural signals input therein, each of the processor elements comprising at least one non-linear oscillator circuit which is designed to bring about the entrainment effect at the first frequency peculiar to the speech of a particular talker; a detection means for detecting the entrained state of each of the processor elements; an extraction means for extracting the aural signals of a particular talker from the aural signals input therein on the basis of the frequency of oscillations of the output signal from each of the processor elements entrained when the means receives the output from the detection means; and an information processing means which is connected to the extraction means and which performs information processing such as word recognition and so on of the aural signals of a particular talker extracted by the extraction means.

In accordance with a preferred form of the present invention, each of the processor elements comprises two non-linear oscillator circuits.

In accordance with a preferred form of the present invention, talker recognition is so set that entrainment of the corresponding processor element takes place at the average pitch frequency of a particular talker.

DESCRIPTION OF THE DRAWINGS

Fig. 1 is a block diagram of the basic configuration of a speech processing apparatus in accordance with the present invention;

Fig. 2 is a drawing of van der Pol-type non-linear oscillator circuits forming each processor element;

Fig. 3 is an explanatory view of the wiring in the case where each processor element comprises two van der Pol circuits;

Fig. 4 is a detailed explanatory view of the configuration of a preprocessing unit;

Fig. 5 is an explanatory view of the connection between a storage block, a regulation modifier and an information generating block;

Fig. 6 is an explanatory view of the connection between a host information processing unit, a modifier, an information generating block and a storage block;

Fig. 7 is an explanatory view of the configuration of a host information processing unit;

Fig. 8 is an explanatory view of another example of the preprocessing unit; and

average pitch frequency detected is extracted from the stored aural signals 401 by the information generating block 305, the signal extracted is an aural signal peculiar to the particular talker.

Non-linear oscillator circuit

The preprocessing unit 2 serves as a central unit of the system in this embodiment. Either of the information generating block 305 or the storage block 306 which serves as a central part comprises a plurality of non-linear oscillator circuits or the like.

In accordance with the understanding of the inventors, the contents of information can be encoded into the phase or frequency of a non-linear oscillator, and the magnification of information can be represented by using the amplitude of the oscillation thereof. In addition, the phase, frequency and amplitude of oscillation can be changed by causing interference between a plurality of oscillators. Causing such interference corresponds to conventional information processing. The interaction between a plurality of non-linear oscillators which are connected to each other causes deviation from the individual intrinsic frequencies and thus mutual excitation, that is "entrainment". In other words, two types of information processing, i.e. the recall of memory performed in the storage block 306 and extraction of the aural signals of a particular talker which is performed in the information generating block 305, are carried out in the preprocessing unit 2. These two types of information processing in the preprocessing unit 2 are performed by using the entrainment taking place owing to the mutual interference between the nonlinear oscillator circuits.

The entrainment is a phenomenon which is similar to resonance and in which all the oscillator circuits make oscillations with the same frequency, amplitude and phase owing to the interference therebetween even if the intrinsic frequencies of the oscillator circuits are not equal to each other. Such entrainment taking place by the interference between the nonlinear oscillators which are coupled with each other is explained in detail in "Entrainment of Two Coupled van der Pol Oscillators by an External Oscillation" (Bio. Cybern. 51, 325-333 (1985)).

It is well known that such a nonlinear oscillator circuit is configured by assembling a van der Pol oscillator circuit using resistor, capacitor, induction coil and negative resistance elements such as a Esaki diode. This embodiment commonly utilizes as a nonlinear oscillator circuit such a van der Pol oscillator circuit as shown in Fig. 2.

In Fig. 2, reference numerals 11a, 12a, 13, 14, 15a, 16 and 17 respectively denote an operational amplifier in which the signs + and - respectively denote the polarities of output and input signals. The resistors 11b, 12b and the capacitors 11c, 12c which are shown in the drawing are applied to the operational amplifiers 11a, 12a, respectively, to form integrators 11, 12. A resistor 15b and a capacitor 15c are applied to the operational amplifier 15a to form a differentiator 15. The resistors shown in the drawing are respectively applied to the other operational amplifiers 13, 14, 16, 17 to form adders. The van der Pol circuit in this embodiment is also provided with multipliers 18, 19. In addition, voltages are respectively input to the operational amplifiers 13, 14, 17 serving as the adders through variable resistors 20 to 22, the variable resistors 20, 21 being interlocked with each other.

The oscillation of this van der Pol oscillator circuit is controlled through an input terminal I in such a manner that the amplitude of oscillation is increased by applying an appropriate positive voltage to the terminal I and it is decreased by applying a negative voltage thereto. A gain controller 23 can be controlled by using the signal input to an input terminal F so that the basic frequency of oscillation of the van der Pol oscillator circuit can be changed. In the oscillator circuit shown in Fig. 2, the basic oscillation thereof is generated by a feedback circuit comprising the operational amplifiers 11, 12, 13, and another part, for example, the multiplier 18, provides the oscillation with nonlinear oscillation characteristics.

As described above, the entrainment is achieved by utilizing interference coupling with another van der Pol oscillator circuit. When the van der Pol oscillator circuit shown in Fig. 2 is coupled with another van der Pol oscillator circuit having the same configuration, the signal input from the other van der Pol oscillator circuit is input in the form of an oscillation wave to each of the terminals A, B shown in Fig. 2, as well as the oscillation wave being output from each of the terminals P, Q shown in the drawing (refer to Fig. 3). When there is no input, the phases of the output P, Q are 90° deviated from each other and when interference input is applied from the other oscillator circuit, this phase difference between output P, Q is changed in correspondence with the relationship between the input and the oscillation wave thereof, as well as the frequency and amplitude being changed.

This embodiment utilizes as a processor element forming each of the storage block 306 and the information generating block 305 an element comprising the two van der Pol nonlinear oscillator circuits

width Δ_{M1} respectively satisfy the above-described equations (1) an (2). This setting will be described below with reference to Fig. 6.

The aural signals 402 from the speech converting block 4 are input to the terminals 610, 611 of each of the processor elements of the storage block 306.

On the other hand, the information generating block 305 also has a plurality of such processor elements 402 as shown in Fig. 3. In the example shown in Fig. 4, q processor elements 402 are provided in the unit 305. The number of processor elements required in the information generating block 305 must be determined depending upon the degree of resolution with which the speech of a particular talker is desired to be extracted. Each of the processor elements 402 of the information generating block 305 also functions as a band pass filter in the same way as the processor elements 403 of the storage block 306. If the processor elements 402 are numbered in turn from the above element and the numerals of the element are denoted by k, the transmission frequency ω_k at which the processor element k functions as a band pass filter is determined so as to have the relationship (3) described below to the basic pitch frequency ω_p of the talker recognized in the storage block 306.

$$\omega_k = k\,\omega_p \quad (3)$$

In other words, in the q processor elements 402a to 402q, their central frequencies ω_{G1}, ω_{G2} ....ω_{Gq} and the band widths Δ_{G1}, Δ_{G2} ... Δ_{Gq} are respectively set so as to satisfy the equations (1) and (2). This setting in the processor elements 402 is described in detail below with reference to Fig. 5.

Each of the storage block 306 and the information generating block 305 has the above described arrangement.

As described above, the processor elements 402 of the information generating block 305 and the processor elements 403 of the storage block 306 are respectively band pass filters having central frequencies which are respectively set to ω_{M1}, ω_{M2} ... ω_{Mr} and ω_{G1}, ω_{G2} ... ω_{Gq}. However, each of these processor elements does not functions simply as a replacement for a conventional known band pass filter, but it efficiently utilizes the characteristics as a processor element comprising nonlinear oscillator circuits. The characteristics include the easiness of modifications of the central frequencies expressed by the equation (1) and the band widths expressed by the equation (2) as well as a high level of selectivity for frequency and responsiveness, as compared with conventional band pass filters.

In the storage block 306, collations of the aural signals 402 with the pitch frequencies previously stored for a plurality of talkers are simultaneously performed for each of the talkers to create an arrangement of the talkers contained in the conversation. That is, the arrangement of talkers contained in conversation can be determined by recognizing the talkers giving speech having the pitch frequencies contained in the conversation expressed by the aural signals 411. The storage of the pitch frequencies in the processor elements 403a to 403r of the storage block 306 is realized by interference oscillation of the processor elements with the basic frequency which is determined by the signals ω_A, ω_B input to the terminal F, as described above with reference to Fig. 3. In other words, the pitch frequencies of the talkers are respectively stored in the forms of the basic frequencies of the processor elements. If the aural signals 411 contain the speech signals of talkers having pitch frequency components ω_2, ω_3 which are close to ω_{M2}, ω_{M3} (i.e., $\omega_2 \simeq \omega_{M2}$ and $\omega_3 \simeq \omega_{M3}$), the processor elements 403a, 403b alone interfere with the input aural signals 411, are activated so as to be entrained and make oscillation with the frequencies ω_2, ω_3, respectively. That is, in the case of conversation of a plurality of talkers, only the processor elements having the frequencies which are set to values close to the average pitch frequencies of the talkers are activated, this activation corresponding to the recall of memory.

The results 501 recalled in the processor elements 403 of the storage block 306 are sent to the processing modifier 303. The processing modifier 303 has the function of detecting the frequencies of the output signals 501 from the processor elements 403, as well as the function of calculating the processing regulation used in the information generating block 305 from the oscillation detected. This processing regulation is defined by the equation (3).

In the information generating block 305, a significant portion, that is, the feature contributing to a particular talker, is extracted from the signals 411 input from the speech converting block 4 in accordance with the processing regulation supplied from the processing regulation modifier 303, and then output as a binary signal to the host information processing unit 3 through the transferrer 307. The binary signal is then subjected to speech processing in the unit 3 in accordance with the demand.

The configuration of talkers can also be recognized by virtue of the host information processing unit 3 based on the information sent from the storage block 306 to the host information processing unit 3 through the transferrer 308.

The information generating block 305 is also capable of adding talkers to be handled and setting parameter data thereof as well as removing talkers.

comparison to the processing unit 3.

The above-described configuration enables the host information processing unit 3 to activate or deactivate any one desired processor element of the storage block 306 or to set/modify the band width and the central frequency thereof.

When a particular one processor element determined by the modifier 309 is activated by the input aural signals 411, and when the pitch frequency ω_p thereof is detected by the modifier 303, the aural signal of the particular talker alone is extracted from the aural signals 411, as described in Fig. 5.

Host Unit

Fig. 7 is a functional block diagram of the processing in the host information processing unit 3 in which speech recognition and talker recognition (talker collation) are mainly performed. One subject of the present invention lies in the processing of the speech signals used for two types of recognition in the preprocessing unit. Since these two types of recognition themselves are already known, they are briefly described below.

The aural signal 412 from the transferrer 307 of the preprocessing unit 2 is a signal containing only the speech of a particular talker. This signal is A/D converted in the transferrer 307 and then input to the processing unit 3. The signal 412 is subjected to cepstrum analysis in 600a in which spectrum estimation is made for the aural signal 412. In such spectrum estimation, the formants are extracted by 600b. The formant frequencies are frequencies at which concentration of energy appears, and it is said that such concentration appears at several particular frequencies which are determined by phonemes. Vowels are characterized by the formant frequencies. The formant frequencies extracted are sent to 601 where pattern matching is conducted. In this pattern matching, speech recognition is performed by DP matching (502a) which is performed for the syllables previously stored in a syllable dictionary and the formant frequencies and by statistical processing (602b) of the results obtained.

A description will now be given of the talker recognition performed in the unit 3.

Although rough talker recognition is carried out in the storage block 306 of the preprocessing unit 2, the talker recognition conducted in the unit 3 is more positive recognition which is carried out using a talker dictionary 605 after the rough talker recognition has been carried out.

In the talker dictionary 605, are stored data with respect to the level of the formant frequency, the band width thereof, the average pitch frequency, the slope and curvature in terms of frequency of the spectral outline and so forth of each of talkers, all of which are previously stored, as well as the time length of words peculiar to each talker and the pattern change with time of the formant frequency thereof.

Application

An application example of the system in the embodiment shown in Fig. 1 is described below with reference to Fig. 8. This application example is configured by adding a switch 801 to the system shown in Fig. 1 so that an information generating section 5 is operated only when the speech of a particular talker is recognized by a storage section 6, and the speech of the particular talker alone is extracted and then sent to the information processing unit 3.

As in the system shown in Fig. 1, a plurality of the processor elements 403 of the storage block 306 comprise one processor element which is activated to the pitch frequency of a particular talker by the modifier 309. When the pitch frequency of the particular talker is detected by the modifier 303, the modifier 303 outputs a signal 802 to the switch 801 so as to close it. In other words, when the switch 801 is opened, the storage block 305 does not operate. In this way, when the switch 801 is turned on, the extraction of only a portion in the aural signals 411 which is also significant from the viewpoint of time by the information generating section 5 enables rapid processing in the host unit 3.

A talker recognition/selector circuit 606 recognizes the talkers by collating the formants extracted by the circuit 600 with the data stored in the dictionary 605. 607 is a r-bit buffer to store the result of talker collation detected by the transferrer 308. Each bit represents whether or not the corresponding comparator of the transferrer 308 has detected that the corresponding processor element of the storage block 306 has been entrained. The circuit 606 compares the result stored in the buffer 607 with the result of talker recognition based on the formant matching operation. Thereby, the talker recognition in the storage block 306 can be confirmed within the processing unit 3.

A r-bit buffer 608 is used to temporarily store the information 409a to 409c.

Claims

1. A speech processing apparatus having input means for inputting the speech of a plurality of talkers and outputting aural signals, said apparatus being characterized by comprising:
a plurality of speech collation processor elements for performing speech collation of said aural signals input therein, each of said processor elements comprising at least one nonlinear oscillator circuit which is so set as to be entrained at a first frequency that characterizes the speech of a talker to be specified;
detection means for detecting the entrained state of each of said processor elements; and
extraction means for extracting the aural signal of a particular talker form said aural signals input therein on the basis of the frequency of the signal output from the entrained processor element when it receives the output from said detection means.

2. A speech processing apparatus according to Claim 1, wherein said nonlinear oscillator circuit is a van der Pol oscillator circuit.

3. A speech processing apparatus according to Claim 1, wherein said first frequency characterizing said speech of said particular talker is the average pitch frequency contained in said speech.

4. A speech processing apparatus according to Claim 1, wherein said speech collation processor element comprises two nonlinear oscillator circuits each of which contains an oscillation control circuit for setting the basic frequency of the oscillation thereof, the difference between the basic frequencies of oscillation of said two nonlinear oscillator circuits and the average frequency thereof respectively corresponding to the band width and the central frequency within a range where said entrainment takes place.

5. A speech processing apparatus according to Claim 1, wherein said extraction means comprises a plurality of speech extraction processor elements for extracting the aural signal fo a particular talker from said aural signals input therein, each of said speech extraction processor elements comprising at least one nonlinear oscillator circuit which is so set as to be entrained at a frequency of integral multiple of said first frequency.

6. A speech processing apparatus according to Claim 1, wherein each of said speech extraction processor element comprises two nonlinear oscillator circuits each of which comprises an oscillation control circuit for setting the basic frequency of the oscillation thereof, the difference between said basic frequencies of said nonlinear oscillator circuits and the average frequency respectively corresponding to the band width and the central frequency in a range where said entrainment takes place.

7. A speech processing apparatus according to Claim 1 further comprising modification means for modifying each of said first frequencies which is so set that each of said speech collation processor elements is entrained.

8. A speech processing apparatus according to Claim 1 further comprising means for inhibiting any entrainment of each of said speech collation processor elements.

9. A speech processing apparatus having input means for inputting a speech and outputting an aural signals of a plurality of specified talkers, for specifying at least one talker from the speech thereof, said apparatus being characterized by comprising:
a plurality of speech collation processor elements for performing speech collation of said aural signals input therein, each of said processor elements comprising at least one nonlinear oscillator circuit which is so set as to be entrained at a first frequency that characterizes the speech of a talker to be specified; and
detection means for detecting the entrained state of each of said processor elements.

10. A speech processing apparatus according to Claim 9, wherein said nonlinear oscillator circuit is a van der Pol oscillator circuit.

11. A speech processing apparatus according to Claim 9, wherein said second frequency characterizing said speech of said talker is an average pitch frequency contained in said speech.

12. A speech processing apparatus according to Claim 9, wherein each of said speech collation processor elements comprises two nonlinear oscillator circuits each of which contains an oscillator control circuit for setting the basic frequency of the oscillation thereof, the difference between said basic frequencies of oscillation of said nonlinear oscillator circuits and the average value thereof respectively corresponding to the band width and the central frequency within the range where said entrainment takes place.

13. A speech processing system having input means for inputting speech of a plurality of talkers and outputting the aural signals thereof, said apparatus being characterized by:
a plurality of speech collation processor elements for performing speech collation of said aural signals input therein, each of said processor element comprising at least one nonlinear oscillator circuit which is so set as to create entrainment at a third frequency that characterizes the speech of a talker to be specified;
detection means for detecting the entrained state of each of said processor elements;

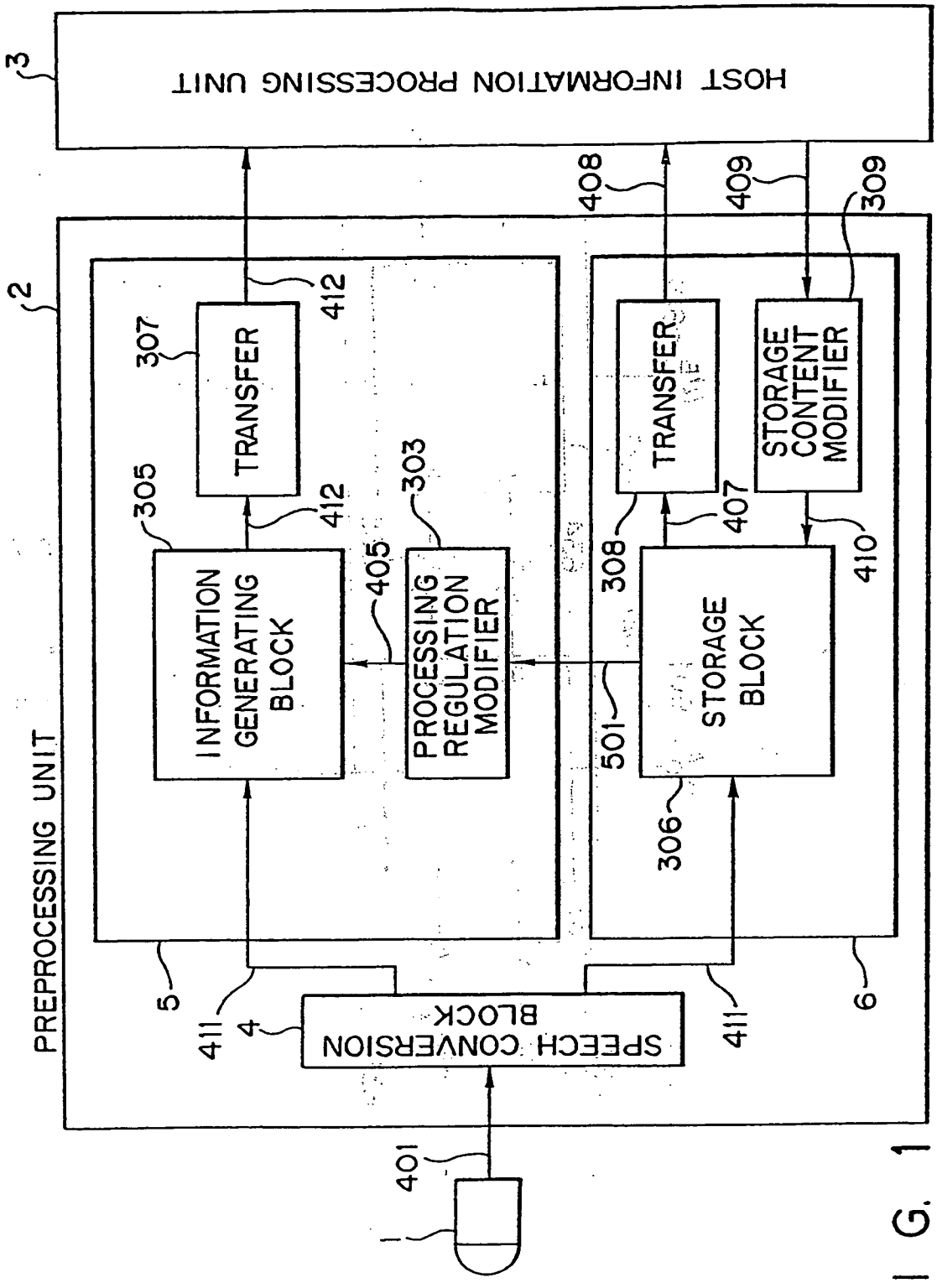
HOST INFORMATION PROCESSING UNIT
3
PREPROCESSING UNIT
2
TRANSFER
307
412
INFORMATION GENERATING BLOCK
305
PROCESSING REGULATION MODIFIER
303
405
TRANSFER
STORAGE CONTENT MODIFIER
STORAGE BLOCK
308
407
410
501
306
309
408
409
SPEECH CONVERSION BLOCK
4
5
6
411
401

FIG. 1

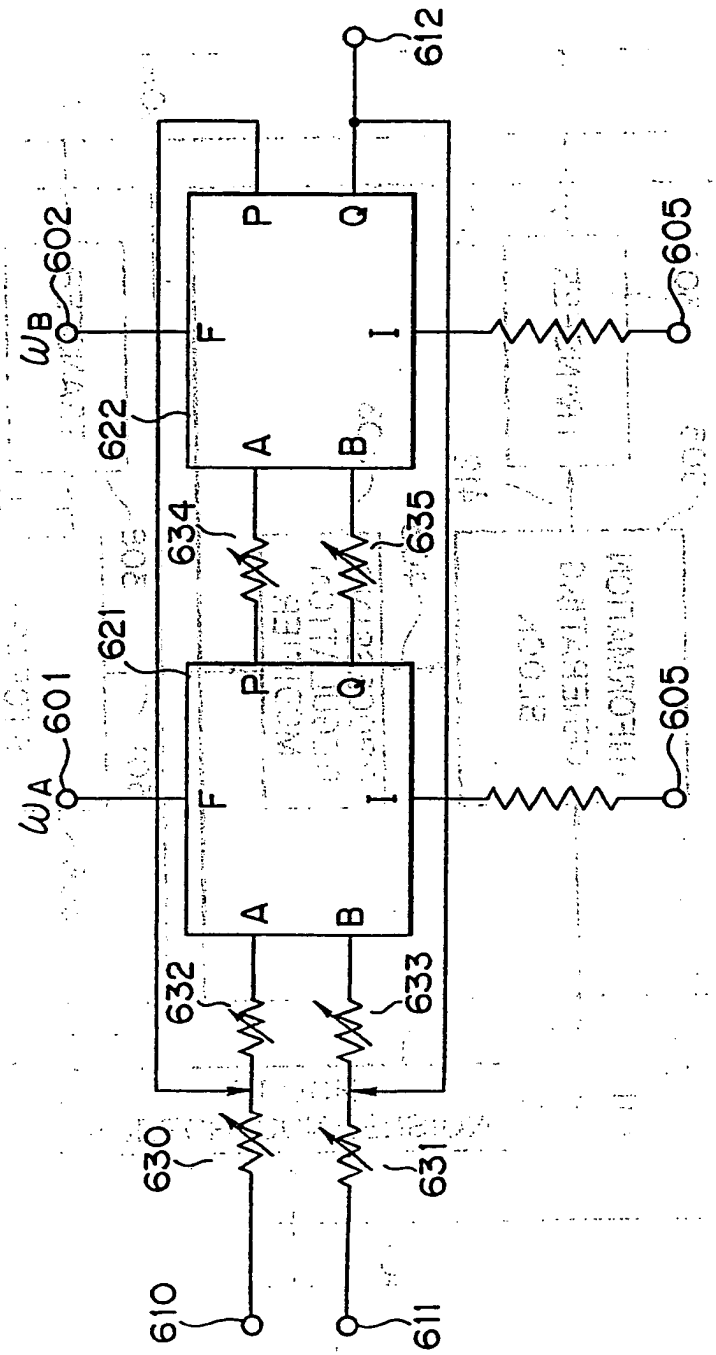
612
602
ω_B
F
P
Q
I
A
B
622
605
634
635
621
601
ω_A
F
P
Q
I
A
B
605
632
633
630
631
610
611

FIG. 3

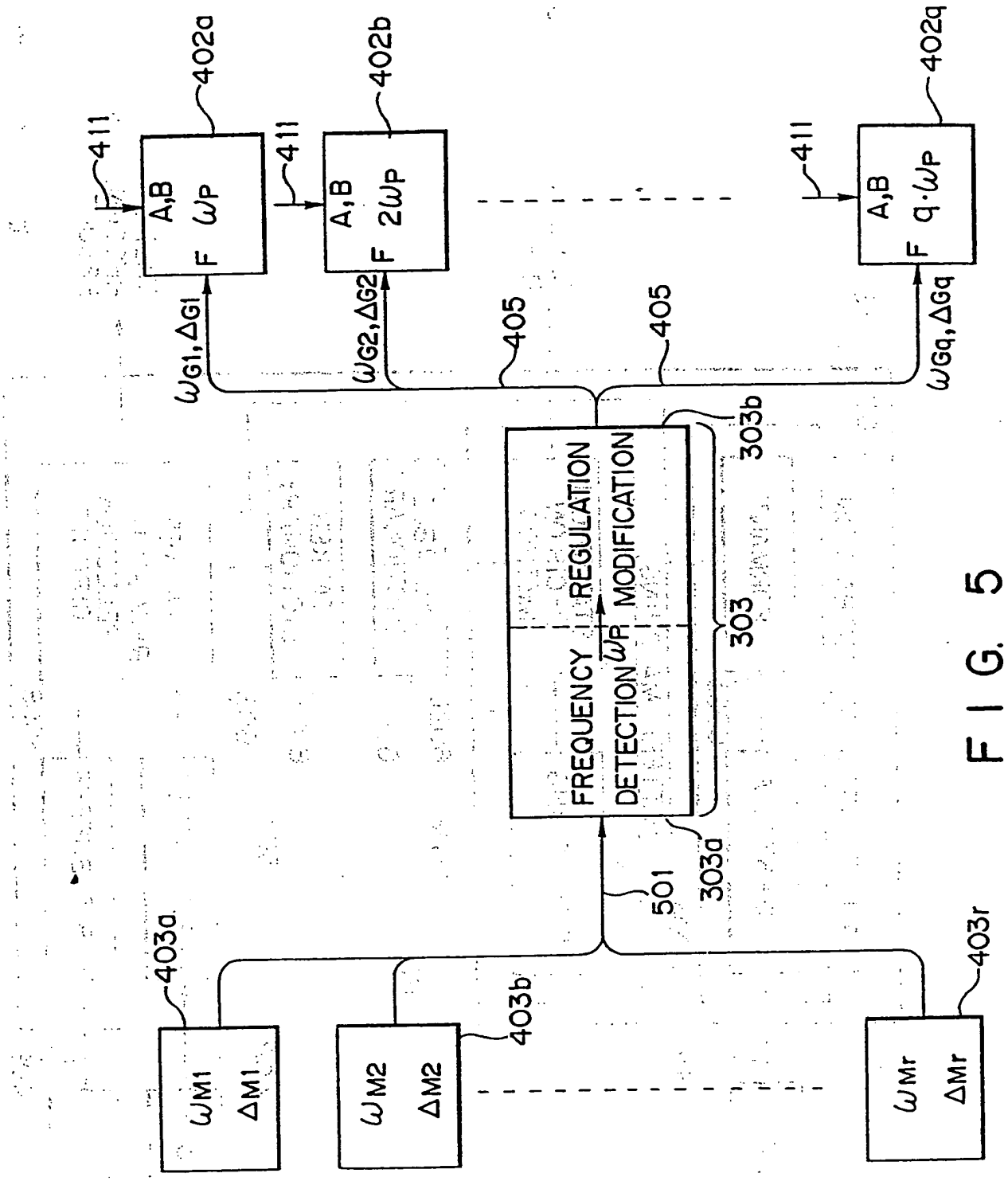
402a
402b
402q
411
411
411
A,B
F ωP
A,B
F 2ωP
A,B
F q·ωP
ωG1, ΔG1
ωG2, ΔG2
ωGq, ΔGq
405
405
303b
303
303a
FREQUENCY DETECTION
REGULATION MODIFICATION
ωP
501
403a
403b
403r
ωM1 ΔM1
ωM2 ΔM2
ωMr ΔMr

FIG. 5

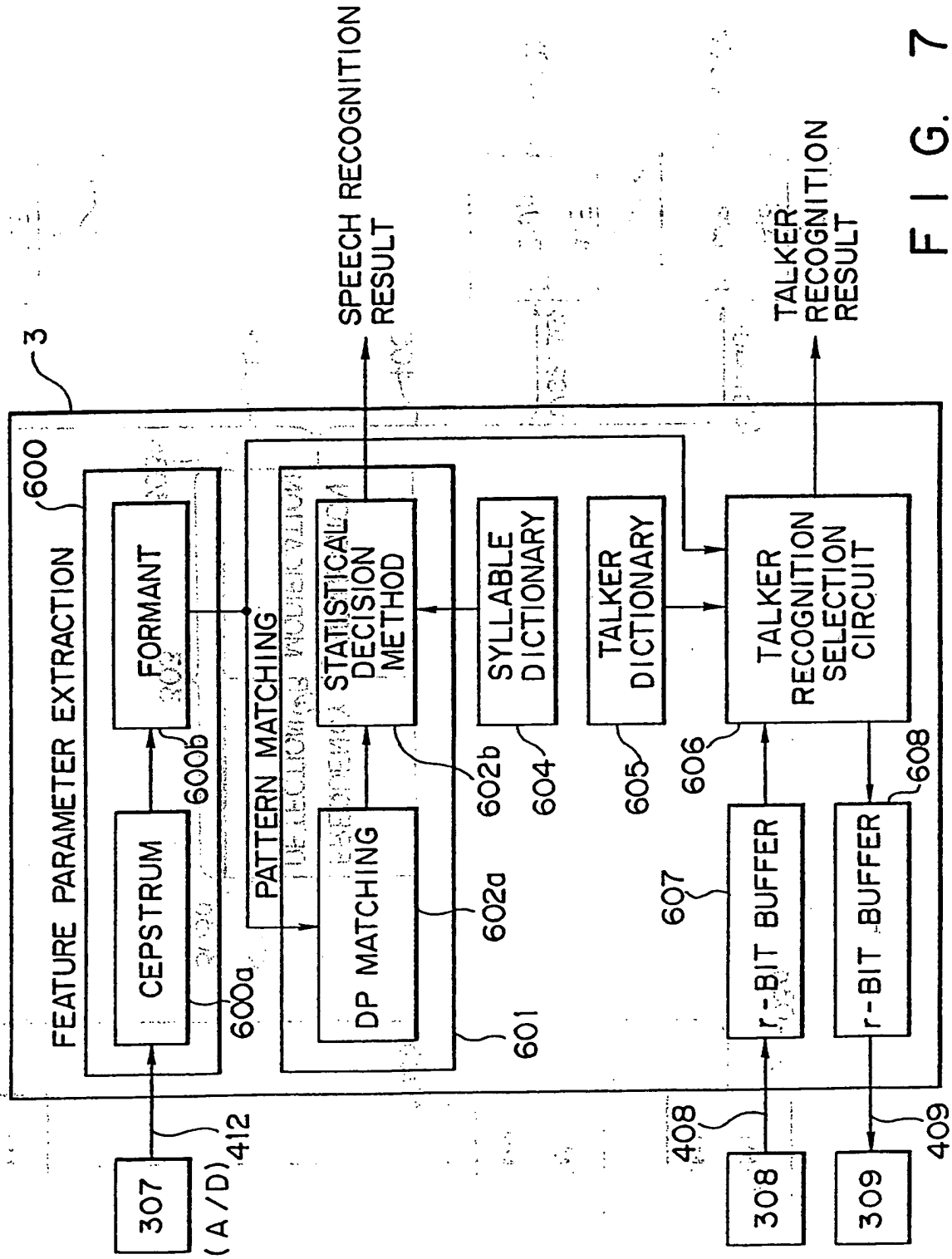
3
600
FEATURE PARAMETER EXTRACTION
CEPSTRUM
600a
FORMANT
600b
PATTERN MATCHING
DP MATCHING
602a
STATISTICAL DECISION METHOD
602b
601
SYLLABLE DICTIONARY
604
TALKER DICTIONARY
605
TALKER RECOGNITION SELECTION CIRCUIT
606
r-BIT BUFFER
607
r-BIT BUFFER
608
SPEECH RECOGNITION RESULT
TALKER RECOGNITION RESULT
307
(A/D)
412
408
409
308
309

FIG. 7

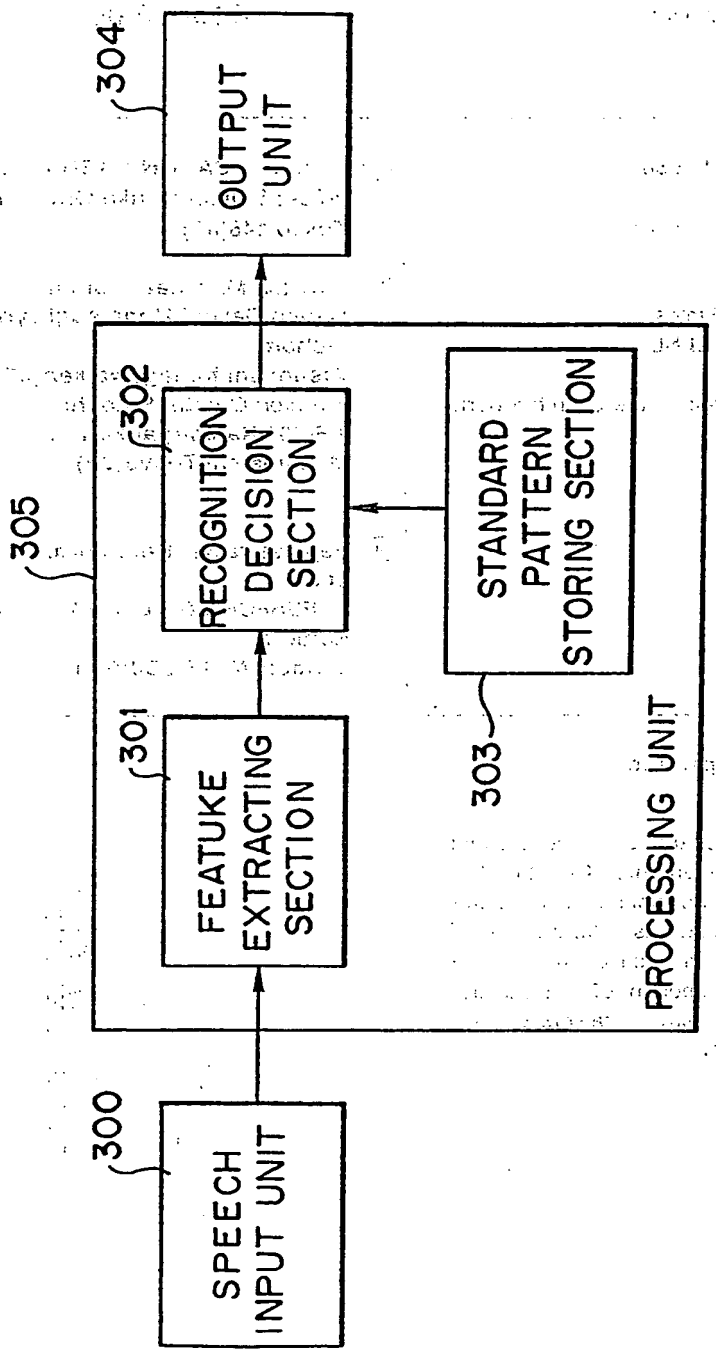
300
SPEECH INPUT UNIT
301
FEATUKE EXTRACTING SECTION
302
RECOGNITION DECISION SECTION
303
STANDARD PATTERN STORING SECTION
304
OUTPUT UNIT
305
PROCESSING UNIT

FIG. 9

(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) Publication number: **0 339 891**
A3

(12) EUROPEAN PATENT APPLICATION

(21) Application number: 89304017.0

(51) Int. Cl.5: G10L 3/00

(22) Date of filing: 21.04.89

(30) Priority: 23.04.88 JP 101173/88

(43) Date of publication of application:
02.11.89 Bulletin 89/44

(84) Designated Contracting States:
AT BE CH DE FR GB IT LI NL

(88) Date of deferred publication of the search report:
16.08.90 Bulletin 90/33

(71) Applicant: CANON KABUSHIKI KAISHA
3-30-2 Shimomaruko Ohta-ku
Tokyo 146(JP)

(72) Inventor: Miyamae, Koichi
Canon Daiichi Honatsugi-ryo 6-29 Mizuhiki 2-chome
Atsugi-shi Kanagawa-ken(JP)
Inventor: Omata, Satoshi
1-5-101 Narusegaoka 1-chome
Machida-shi Tokyo(JP)

(74) Representative: Beresford, Keith Denis Lewis et al
BERESFORD & Co. 2-5 Warwick Court High Holborn
London WC1R 5DJ(GB)

(54) Speech processing apparatus.

(57) A speech processing apparatus of the present invention enables processor elements (403a to 403r) each comprising at least one nonlinear oscillator circuit (621) to be used as band pass filters by using the entrainment taking place in each of the processor elements, whereby the speech of a particular talker in the speech of a plurality of talkers can be recognized.

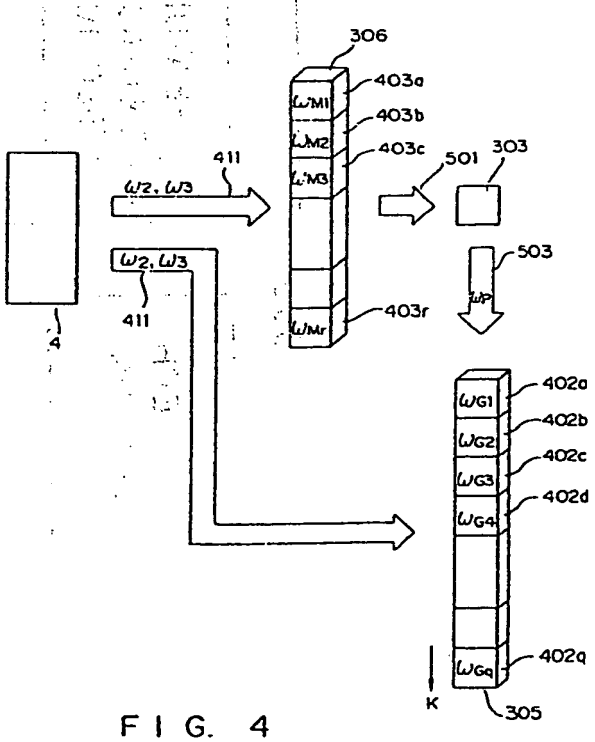

FIG. 4

EP 0 339 891 A3